

REMARKS

In the Office Action dated September 29, 2009, the Examiner rejected claims 1-3, 5-12, 14-17, 19-27, 29-31, 33-38 and 40-51. Claims 1 and 27 are amended. Claims 1-3, 5-12, 14-17, 19-27, 29-31, 33-38 and 40-51 remain pending upon entry of this amendment.

Summary of the Rejections

Claims 1-3, 5-12, 14-17, 19-27, 29-31, 33-38 and 40-51 are rejected under 35 U.S.C. §103(a) as being unpatentable over Bozdagi (USPN 6,647,535) in view of Yang (USPN 6,301,586).

These rejections are now traversed.

Response to All Rejections under 35 U.S.C. § 103(a)

The rejections are addressed by reference to the independent claims.

Independent Claim 1

Applicants' amended claim 1 recites, in part:

A system for printing comprising:
a user interface for receiving instructions from a user for **controlling segmentation of audio or video time-based media content for printing based on one or more features** within the audio or video time-based media content, the features including any of speech recognition, optical character recognition, facial recognition, speaker detection, facial detection and event detection, and for generation of a printable representation of the media content, the user interface comprising a content selection field displaying a graphical representation of the audio or video time-based media content and the instructions from the user comprising selection of a segment of the graphical representation of the audio or video time-based media content;

a media analysis module communicatively coupled to the user interface, the media analysis module **analyzing the features of the audio or video time-based media content to extract** the segment of the audio or video time-based media content selected from the graphical representation based at least in part on the instructions received from the user in the user interface; . . . (emphasis added)

The combination of Bozdagi and Yang fails to teach or suggest “a user interface for receiving instructions from a user for controlling segmentation of audio or video time-based media content for printing based on one or more features within the audio or video time-based media content, the features including any of speech recognition, optical character recognition, facial recognition, speaker detection, facial detection and event detection, and for generation of a printable representation of the media content” and “analyzing the features of the audio or video time-based media content to extract the segment of the audio or video time-based media content.”

Bozdagi describes a system for parsing a video into representative frames to avoid storage and bandwidth problems.¹ The frames are presented to a user who can browse through the segments by reviewing the representative frames.²

Bozdagi does not teach or suggest controlling segmentation of audio or video time-based media content for printing where the features include any of speech recognition, optical character recognition, facial recognition, speaker detection, facial detection and event detection. The comparing process described in Bozdagi is used to select a key frame for each segment and not segmentation as recited in claim 1. Segmentation describes the process of

¹ Bozdagi, column 2, lines 4-14.

² Bozdagi, column 2, lines 15-18.

segmenting multiple frames. As a result, Bozdagi fails to teach or suggest controlling segmentation of audio or video time-based media content.

In addition, Bozdagi also fails to teach or suggest the features including any of speech recognition, optical character recognition, facial recognition, speaker detection, facial detection and event detection. Bozdagi describes two ways for selecting key frames. First, a frame is selected from the video by extracting command data that is embedded in the closed-caption portion of the data.³ Because the key frames are selected by a programmer that determines the commands and embeds command data into the closed-caption portion of the data, there is nothing to teach or suggest selection of a frame based on the features of the audio or video time-based media content.

In the second method disclosed in Bozdagi, a frame is selected by computing the difference between two consecutive frames on a pixel-by-pixel basis.⁴ Computing the difference between two consecutive frames is not equivalent to using speech recognition, optical character recognition, facial recognition, speaker detection, facial detection or event detection to control segmentation of the video.

Bozdagi also fails to teach or suggest a media analysis module for analyzing the features of the audio or video time-based media content to extract the segment of the audio or video time-based media content where the features include any of speech recognition, optical character recognition, facial recognition, speaker detection, facial detection and event detection. The Examiner asserts on page 9 with reference to claim 29 that because Bozdagi teaches that “a real time video system with a speaker can be represented (column 6, lines 47-

³ Bozdagi, column 4, lines 25-36.

⁴ Bozdagi, column 4, lines 48-50.

65) [it] would have been obvious that performing speech recognition on the video time-based media data was well known in the art.” Applicants reproduce this section of Bozdagi below:

However, command information, such as closed-caption information containing special characters, or text strings, can be embedded in a portion of the multimedia image data signal to indicate, or supplement, a representative or significant image. For example, FIG. 4 illustrates the representative frames and text strings **122** that were derived from an exemplary multimedia image data signal containing command information.

For example, special characters in the command data can indicate representative images, change in speakers, or additional data to be displayed, for example, with the representative image.

With closed-caption data, a change in the speaker can be represented, for example, by the special character string “>>” during production. Thus, for the exemplary commercial segment shown in FIG. 4, this character string acts as the command indicating, for each occurrence, that a new frame and text string **122** are to be captured.

At most Bozdagi discloses programming a data stream that is associated with the video to signify when the video shows a change in the speaker and extracting text strings from the programmed data stream. Bozdagi is silent with regard to speech recognition.

In addition, Applicants respectfully point out that while the concept of speech recognition may be known in the art, the claim language recites that the features are analyzed “to extract the segment of the audio or video time-based media content.” Nowhere in Bozdagi is there the teaching or suggestion of analyzing features, such as speech recognition, to extract the segment of the audio or video time-based media content. In determining the differences between the prior art and the claims, the question under 35 U.S.C. 103 is not whether the differences themselves would have been obvious, but whether the claimed invention as a whole would have been obvious. *Stratoflex, Inc. v. Aeroquip Corp.*, 713 F.2d 1530 (Fed. Cir. 1983).

The Examiner further asserts on page 10 with regard to claim 30 that because Bozdagi teaches that “a real time video system with closed-caption information containing special character[s] . . . [i]t would have been obvious that performing the optical character recognition on the audio (or video-time based) media was well known in the art.” Applicants disagree. In addition to the fact that the Examiner again ignores that the optical character recognition is one of the features being analyzed to extract the segment of the audio or video time-based media content, Applicants also disagree that using optical character recognition on audio or video-time based media content is a concept that is well known in the art. “[R]ejections on obviousness grounds cannot be sustained by mere conclusory statements; instead, there must be some articulated reasoning with some rational underpinning to support the legal conclusion of obviousness.” *In re Kahn*, 441 F.3d 977, 988 (Fed. Cir. 2006).

Lastly, on page 10 of the Office Action with reference to claims 31 and 34, the Examiner asserts that because Bozdagi discloses a video system with closed-caption information and a speaker, it would have been obvious that performing facial recognition on the media was well known in the art. First, Applicants fail to understand how closed-caption information and a video with a speaker relate to the features of facial recognition and facial detection. Second, Applicants again submit that the Examiner is using conclusory arguments to argue that the concepts of claimed invention are well-known in the art. The Examiner cannot arbitrarily combine references and state that teachings are known to those skilled in the art to establish the elements of the claimed invention where there is no teaching or suggestion in the prior art to do so. Such arguments are hindsight construction of obviousness for the claimed invention, which the courts have repeatedly stated is not permissible.

Yang fails to remedy the deficiencies of Bozdagi. Yang discloses a system for managing multimedia clips.⁵ A user can select an image for the folder containing multimedia clips.⁶ For example, a user selects pictures of jewelry for a folder that contains images of a jewelry inventory.⁷ The only discussion of audio or video clips is a brief mention in column 4, lines 45-46 that a network interface is used to access video clips stored on cameras. Yang fails to even discuss segmentation of media content, let alone segmentation based on features including any of speech recognition, optical character recognition, facial recognition, speaker detection, facial detection and event detection. As a result, the combination of Bozdagi and Yang fail to teach or suggest the above-recited features of claim 1.

Claims 2, 3, 5-12, 14-17, 19-26 and 49-51 depend upon claim 1. As a result, they are patentable for at least the same reasons as claim 1.

Independent Claim 27

Applicants' amended claim 27 recites, in part:

A method for printing, the method comprising:

displaying a print dialog driver box to a user wherein the print dialog driver box comprises a user interface for receiving instructions from the user for **controlling segmentation of audio or video time-based media content for printing based on one or more features within the audio or video time-based media content, the features including any of speech recognition, optical character recognition, facial recognition, speaker detection, facial detection and event detection**, and for generation of a printable representation of the audio or video time-based media content, the user interface comprising a content selection field displaying a graphical representation of the audio or video time-based media content and the instructions from the

⁵ Yang, Abstract.

⁶ Yang, column 1, lines 59-64.

⁷ Yang, column 2, lines 2-6.

user comprising selection of a segment of the graphical representation of the audio or video time-based media content. . .

analyzing the features of the audio or video time-based media content to extract the segment of the audio or video time-based media content selected from the graphical representation based at least in part on the instructions received from the user in the print dialog driver box; . . . (emphasis added)

The cited references do not disclose or suggest the above-quoted elements of Applicants' claim 27 for at least the same reasons as those described for claim 1. Accordingly, claim 27 is patentable over the cited references. Claims 29-31, 33-38 and 40-48 incorporate the limitations of claim 27, and are therefore patentable over the cited references for at least the same reasons as claim 27.

Respectfully submitted,
JONATHAN J. HULL, ET AL.

Dated: June 1, 2010

/Elizabeth Ruzich/
Elizabeth Ruzich, Reg. No. 54,416
Attorney for Applicants
PATENT LAW WORKS LLP
165 South Main Street, Second Floor
Salt Lake City, UT 84111
Tel.: (801) 258-9824
Fax: (801) 355-0160
Email: eruzich@patentlawworks.net